Supplementary Information Appendix

Evolution of the 3-hydroxypropionate bi-cycle and recent transfer of anoxygenic photosynthesis into the Chloroflexi

Authors: Patrick M. Shih[a,b,1], Lewis M. Ward[c], Woodward W. Fischer[c,1]

Affilliations:
[a] Joint BioEnergy Institute, 5885 Hollis St, Emeryville, CA 94608, USA.
[b] Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, One Cyclotron Rd, Berkeley, CA 94720, USA.
[c] Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125, USA.

[1] To whom correspondence should be addressed, E-mail: pmshih@gmail.com or wfischer@caltech.edu

Supplementary Information - Table of Contents

Supplementary Text …….. 3-4
Supplementary References ……..…….. 5-6
Supplementary Figures ……..…….. 7-14
Supplementary Table ………..…… 15-16

Supplemental Text:

Aerobic respiration predates phototrophy in Chloroflexales:

While our results indicate that horizontal gene transfer was the mode for the evolution of phototrophy relatively recently in the evolutionary history of the Chloroflexi, what metabolism(s) characterized the host cells prior to the transfer event?

On the basis of comparative biology, it appears that many—if not all—extant anoxygenic phototrophs are derived within groups that have basal members with aerobic metabolism; and from these observations it was hypothesized that it was somewhat easy for facultative aerobes to pick up phototrophy because of the modular structure of the high-potential electron transport chains that are common to both phototrophy and aerobic respiration (and denitrification)[1]. This also presents a logical juxtaposition that anoxygenic phototrophy in these groups will postdate the rise of oxygen (*ca*. 2.3 Ga) and evolution of widespread aerobic respiration[1,2].

The history of aerobic respiration in the Chloroflexi is richer than previously appreciated, with most lineages containing respiration genes, even if the cultured isolates were originally characterized as obligate anaerobes[3-5]. In order to determine the evolutionary history of aerobic respiration genes, it is useful to compare the phylogenies of proteins to that of the organisms themselves because incongruence between these phylogenies can be indicative of horizontal transfer of genes[6]. Sequences of respiration proteins were identified locally with BLAST+[7], aligned with MUSCLE[8], and alignments manually curated in Jalview[9]. Phylogenetic trees were calculated using RAxML[10] on the Cipres science gateway[11]. Trees were visualized with Seaview[12].

Evolutionary relationships illustrate that O_2 reductases and Complex III/Alternative Complex III are exchanged frequently, and appear to have been acquired independently in in each class (with the exception of the anaerobic Dehalococcoidetes)(Fig. S3). In particular, it appears that Alternative Complex III and a high-affinity B-family heme copper oxidoreductase (HCO) were acquired along with phototrophy at the base of the phototrophic Chloroflexales. Within the total group Chloroflexia class (Chloroflexales+Herpetosiphonales+Kallotenuales), however, the presence of aerobic respiration via a low-O_2-affinity A family HCO appears to be a synapomorphy. A congruent history for this protein is retained between in the lineages Kallotenue+Herpetosiphon+Kouleothrix+Roseiflexus, though it appears a loss has occurred in the branch leading toward Chlorothrix, Oscillochloris, and Chloroflexus, followed by replacement in the Oscillochloris+Chloroflexus lineage. Therefore, aerobic respiration using an A family HCO reductase was acquired in basal members of the Chloroflexia, before the acquisition of photosynthesis in Chloroflexales. This supports the idea that anoxygenic phototrophy was recruited into an ancestral lineage capable of aerobic respiration[1], and provides independent support for our molecular clock results that indicate the acquisition of phototrophy in Chloroflexi after the evolution of oxygenic photosynthesis in Cyanobacteria and the rise of atmospheric oxygen at *ca*. 2.3 Ga (Fig.1; Table 1).

Carbon isotope mass balance analysis:

Carbon fixation metabolisms typically exhibit a kinetic isotope effect leaving organic carbon depleted in ^{13}C relative to dissolved inorganic carbon (DIC), with a characteristic fractionation from each carbon fixation pathway[13]. The 3HP bi-cycle imparts a fractionation as DIC is fixed into organic carbon relative to other carbon fixation pathways (~-13 ‰ compared to ~-25 ‰ for the Calvin Cycle[14]).

If phototrophic Chloroflexi using the 3HP bi-cycle were an important contributor to carbon fixation, one might expect to see this recorded in the C isotope ratios of preserved sedimentary organic matter in the geologic record. However, never in Earth history are bulk sedimentary carbon isotopes so heavy as to be consistent with a dominant 3HP source. Rather, prior to the rise of oxygen and during Archean time (>2.5 billion years ago) the opposite is true, with kerogens recording anomalously low $\delta^{13}C$ of organic carbon of ~-35 to -50‰, potentially indicating a substantial amount of carbon fixation via the reductive acetyl-CoA pathway[15]. One can use isotope mass balance calculations to estimate what proportion of production could be due to the 3HP bi-cycle on the early Earth, constrained by C isotope ratio data.

A simple isotope mass balance calculation can determine the maximum allowable contribution of 3HP to carbon fixation given the observed bulk difference in $\delta^{13}C$ of buried organic carbon and carbonate ($\Delta^{13}C_{org/carb}$) and the average isotopic difference between DIC and organic carbon due to carbon fixation pathways (ε). The bulk difference in $\delta^{13}C$ between carbonate and sedimentary organic carbon, $\Delta^{13}C_{org/carb,}$ is a function of the fractionation between dissolved inorganic carbon and organic carbon imparted by carbon fixation; if multiple carbon fixation pathways are producing organic carbon, the bulk $\Delta^{13}C_{org/carb}$ can be calculated as a simple isotope mass balance mixing calculation of the form $\Delta^{13}C_{org/carb} = fA+(1-f)\ \varepsilon$. f is the fraction of organic carbon fixed via 3HP, A is the fractionation between DIC and organic carbon imparted by 3HP, ε is the average fractionation associated with other active carbon fixation pathways. The average value of A, the fractionation imparted by 3HP, is ~-13‰[14], while the value of ε varies depending on the relative contribution of carbon fixation pathways such as the reductive acetyl CoA pathway (with large fractionations, ~-25 to -69‰, average for hydrogenotrophic methanogens ~-47‰) and the Calvin Cycle (smaller fractionations ~-22 to -30‰, average ~-27‰)[13,14,16-18]. Here, we plotted a contour map of the value of f for given values of $\delta^{13}C_{kerogen}$ and ε . f values of 0 indicate regions that do not have a solution for which f is positive, reflecting ε values too low to produce that value of $\delta^{13}C_{org}$. Given typical values of $\Delta^{13}C_{org/carb,}$ [19-21] and expectations of average fractionations similar to that of the Calvin Cycle today, no intervals in Earth history are consistent with 3HP as a dominant driver of carbon fixation. For the most conservative scenario during Archean time, assuming organic carbon fixed via the reductive acetyl-CoA pathway with an fractionation of ~-47‰ (the average fractionations imparted by hydrogenotrophic methanogens[18]) and a maximal fractionation from 3HP of -13‰, a bulk kerogen $\delta^{13}C$ -50‰ could reflect no more than about 5% of organic carbon fixed via 3HP.

References:

1 Fischer, W., Hemp, J. & Johnson, J. Evolution of oxygenic photosynthesis. *Annu Rev Earth Planet Sci* **44**, 647-683 (2016).

2 Soo, R. M. *et al.* An Expanded Genomic Representation of the Phylum Cyanobacteria. *Genome Biol Evol* **6**, 1031-1045 (2014).

3 Pace, L. A., Hemp, J., Ward, L. M. & Fischer, W. W. Draft Genome of Thermanaerothrix daxensis GNS-1, a Thermophilic Facultative Anaerobe from the Chloroflexi Class Anaerolineae. *Genome Announcements* **3** (2015).

4 Ward, L. M., Hemp, J., Pace, L. A. & Fischer, W. W. Draft Genome Sequence of Leptolinea tardivitalis YMTK-2, a Mesophilic Anaerobe from the Chloroflexi Class Anaerolineae. *Genome Announcements* **3** (2015).

5 Hemp, J., Ward, L. M., Pace, L. A. & Fischer, W. W. Draft Genome Sequence of Levilinea saccharolytica KIBI-1, a Member of the Chloroflexi Class Anaerolineae. *Genome Announcements* **3** (2015).

6 Fitch, W. M. Of urfs and orfs: A primer on how to analyze derived amino acid sequences. *Cell* **50**, 329 (1986).

7 Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).

8 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792-1797 (2004).

9 Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191 (2009).

10 Stamatakis, A. RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics* (2014).

11 Miller, M. A., Pfeiffer, W. & Schwartz, T. in *Gateway Computing Environments Workshop (GCE), 2010.* 1-8.

12 Gouy, M., Guindon, S. & Gascuel, O. SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building. *Mol Biol Evol* **27**, 221-224 (2010).

13 Hayes, J. M. Fractionation of Carbon and Hydrogen Isotopes in Biosynthetic Processes. *Reviews in Mineralogy and Geochemistry* **43**, 225 (2001).

14 Berg, I. A. *et al.* Autotrophic carbon fixation in archaea. *Nat Rev Micro* **8**, 447-460 (2010).

15 Slotznick, S. P. & Fischer, W. W. Examining Archean methanotrophy. *Earth and Planetary Science Letters* **441**, 52-59 (2016).

16 Games, L. M., HayesRobert, J. M. & Gunsalus, P. Methane-producing bacteria: natural fractionations of the stable carbon isotopes. *Geochim Cosmochim Acta* **42**, 1295-1297 (1978).

17 Botz, R., Pokojski, H.-D., Schmitt, M. & Thomm, M. Carbon isotope fractionation during bacterial methanogenesis by CO2 reduction. *Organic Geochemistry* **25**, 255-262 (1996).

18 Blaser, M. B., Dreisbach, L. K. & Conrad, R. Carbon Isotope Fractionation of 11 Acetogenic Strains Grown on H2 and CO2. *Appl Environ Microbiol* **79**, 1787-1794 (2013).

19 Des Marais, D. J., Strauss, H., Summons, R. E. & Hayes, J. M. Carbon isotope evidence for the stepwise oxidation of the Proterozoic environment. *Nature* **359**, 605-609 (1992).
20 Eigenbrode, J. L. & Freeman, K. H. Late Archean rise of aerobic microbial ecosystems. *Proc Natl Acad Sci* **103**, 15759-15764 (2006).
21 Hayes, J. M. & Waldbauer, J. R. The carbon cycle and associated redox processes through time. *Philos Trans R Soc B Biol Sci* **361**, 931 (2006).

Supplementary Figures:

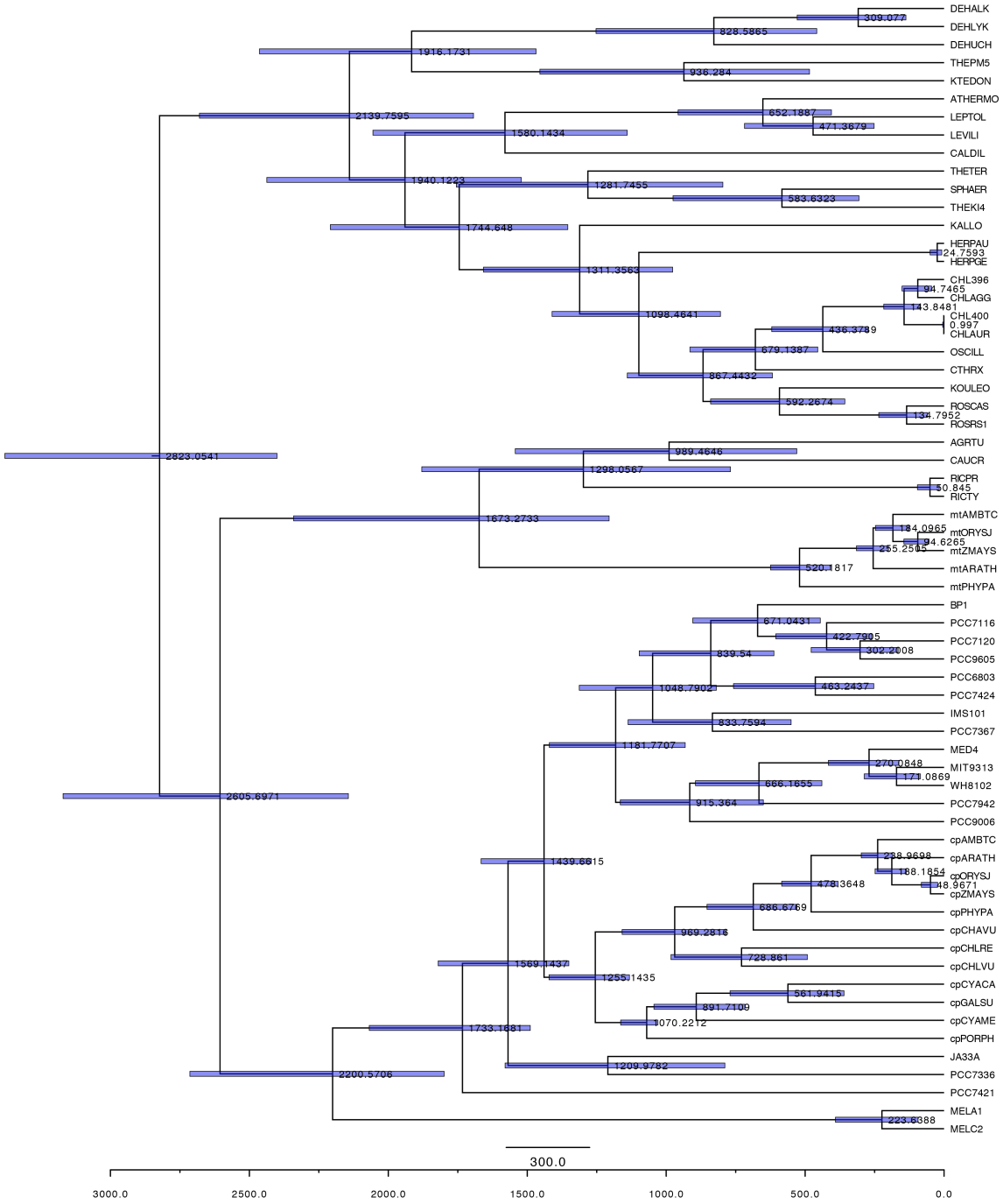

Figure S1. Divergence time estimates from cross-calibrated molecular clock analysis with blue bars making the 95% CI. Taxa abbreviations are summarized in Table S1.

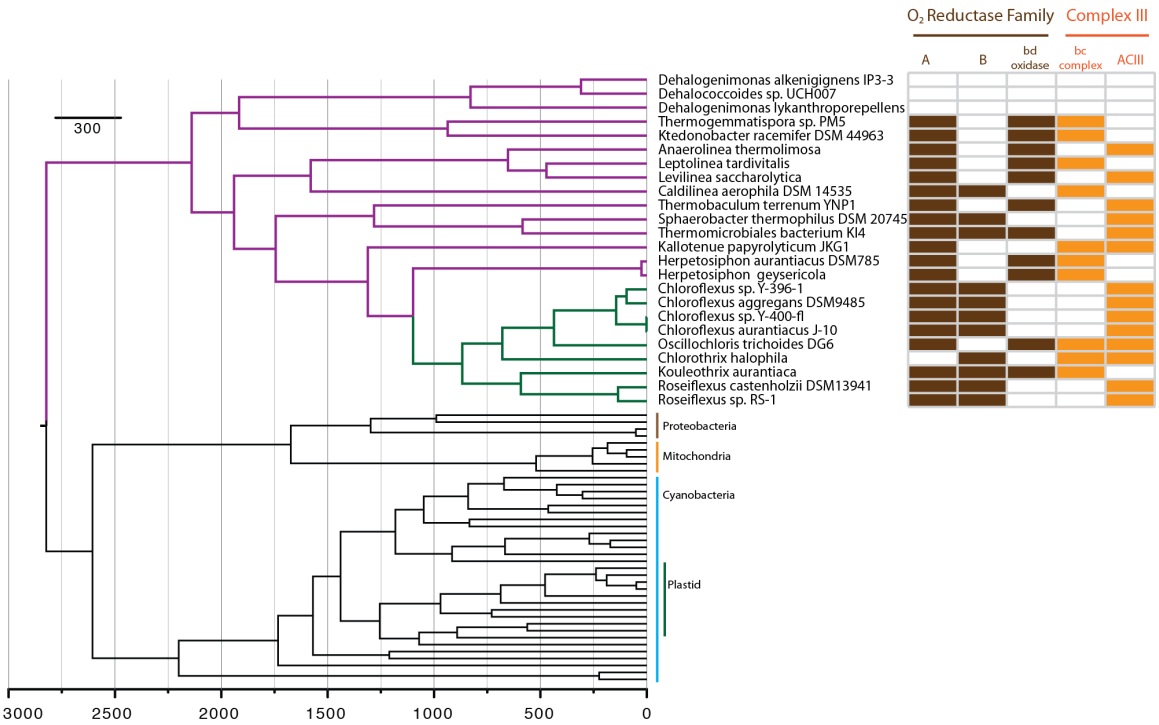

Figure S2. Phylogenetic distribution of aerobic metabolism markers. Members of the A- and B-families of Heme Copper Oxidoreductases (HCOs) are widespread in the Chloroflexia, while *bd* oxidase-type O_2 reductases are also present. The two forms of Complex III (Cytochrome *bd* complex and Alternative Complex III) were also found in various members of the Chloroflexi. All Chloroflexi members have genes that would enable aerobic metabolism, except those from the Dehalococcoidetes, indicating that aerobic metabolism was likely present in the phylum prior to the acquisition of phototrophy in the Chloroflexales.

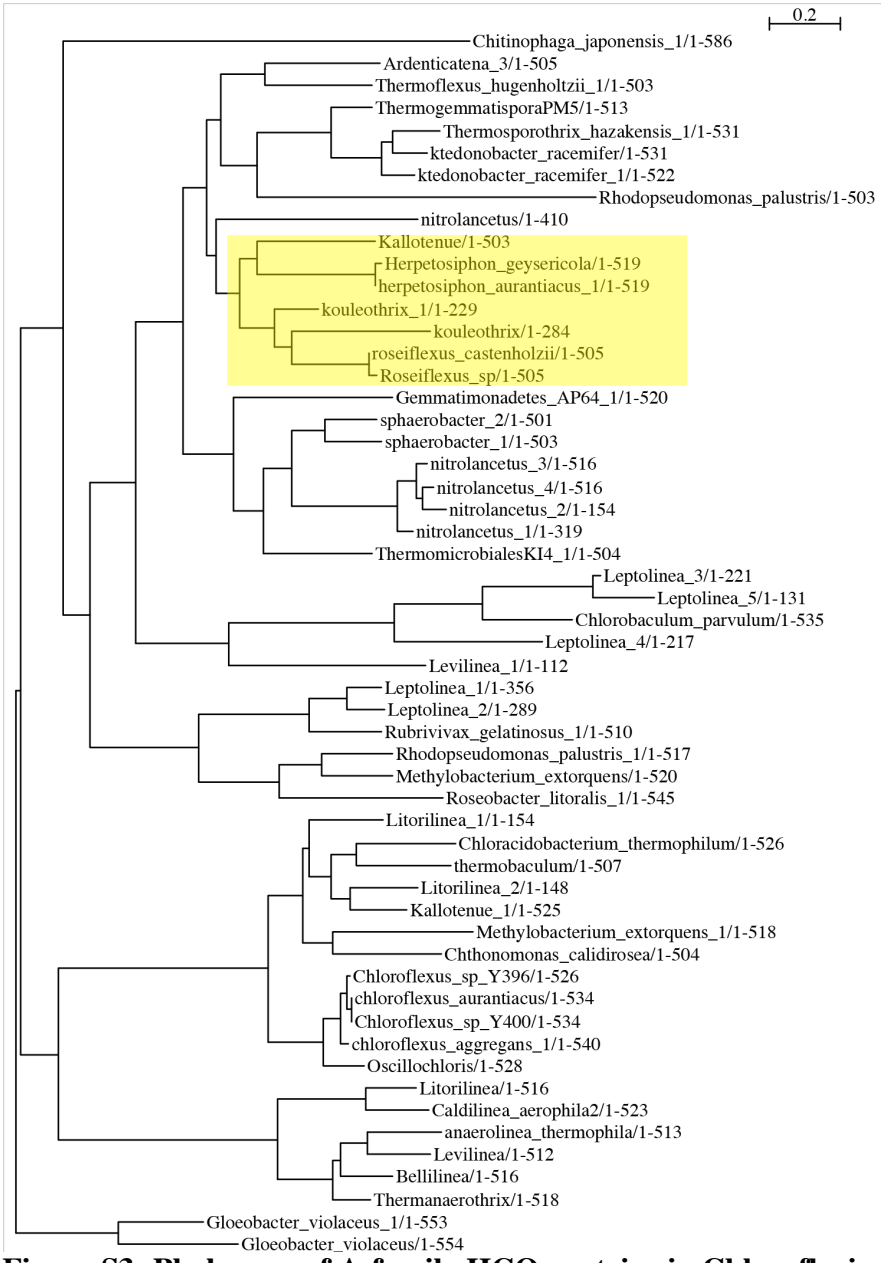

Figure S3: Phylogeny of A-family HCO proteins in Chloroflexi and select outgroups. The overall topology of the tree is discordant from that of 16S rDNA and single-copy phylogenetic protein markers, which is indicative of horizontal gene transfer within the Chloroflexi phylum. However there is congruence within some branches of the tree suggests that aerobic respiration has been vertically inherited within some clades. In particular, The Chloroflexia class appears to have acquired aerobic respiration via transfer of an A-family HCO, before the divergence of *Herpetosiphon* and *Kallotenue* from the phototrophic Chloroflexales, given the congruence of the Herpetosiphon+Kallotenue+Kouleothrix+Roseiflexus clade in both HCO and organismal trees (highlighted in yellow). Loss of the A-family HCO appears to have occurred in Chlorothrix+Oscillochloris+Chloroflexus, followed by reacquisition of a new copy in the Oscillochloris+Chloroflexus lineage.

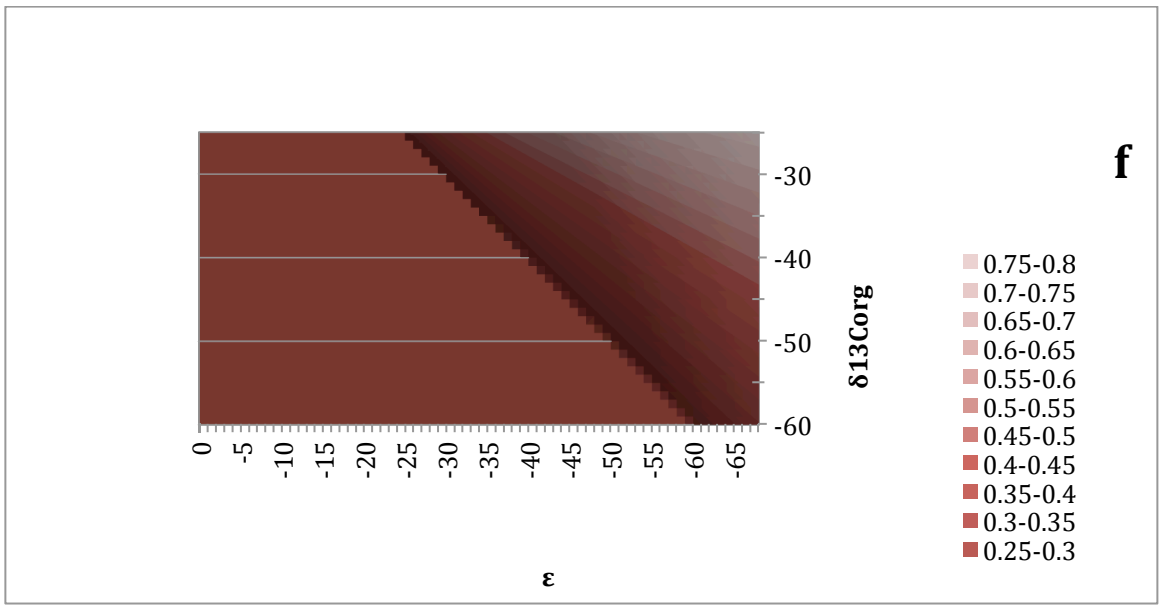

Figure S4. Contours of allowable 3HP mixing ratio to reach average $\Delta^{13}C_{org/carb}$ (difference between sedimentary organic carbon and carbonate $\delta^{13}C$, vertical axis) with average organic carbon fractionation (ε, horizontal axis). $\Delta^{13}C_{org/carb}$ incorporates fractionations from a variety of carbon fixation pathways, each with a characteristic fractionation between DIC and organic carbon. 3HP imparts a characteristically small fractionation between DIC and organic carbon of ~-13 ‰, so if 3HP is ever a significant contributor to productivity it would serve to pull $\delta^{13}C_{org,}$ and therefore $\Delta^{13}C_{org/carb,}$ to heavier values, so if $\Delta^{13}C_{org/carb}$ remains low 3HP must be a small fraction of total productivity. E.g., for 3HP mixing ratio near 1, $\Delta^{13}C_{org/carb}$ would approach -13 ‰ for any ε. For ε near -27‰, typical of the Calvin Cycle, and $\Delta^{13}C_{org/carb}$ of ~-25‰ typical of the Proterozoic when 3HP is predicted to have evolved, the mixing ratio of 3HP must be less than 0.15.

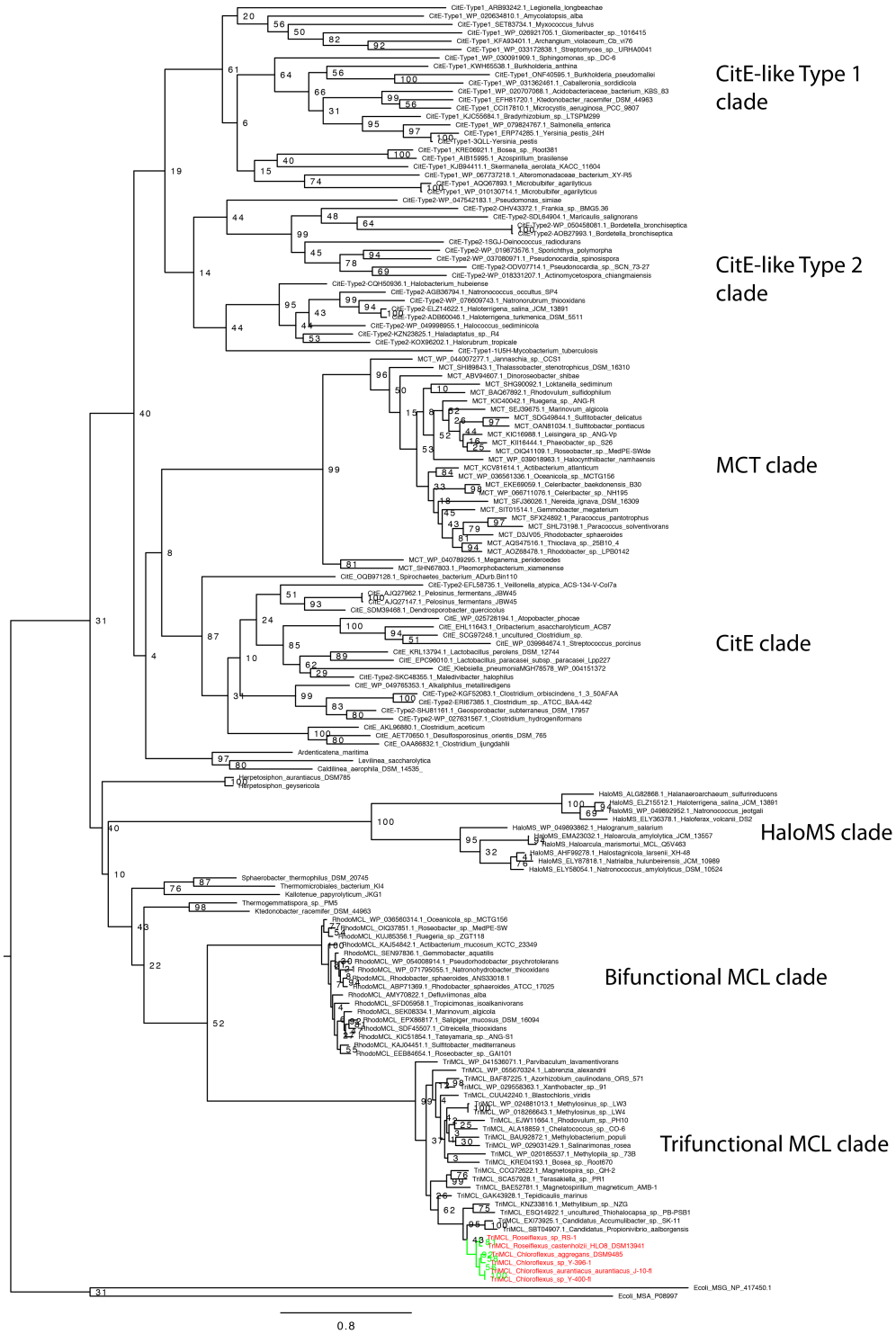

Figure S5. Maximum likelihood phylogeny of MCL/CitE super family. The MCL protein family is only one subclade of the larger CitE superfamily. Chloroflexales trifunctional MCL branches are highlighted in green, and taxa in red. The derived placement of the Chloroflexales taxa demonstrates suggests the late horizontal gene transfer of this key enzyme involved in the 3-hydroxypropionate bi-cycle. Tree is rooted to E. coli malate synthases MSA and MSG.

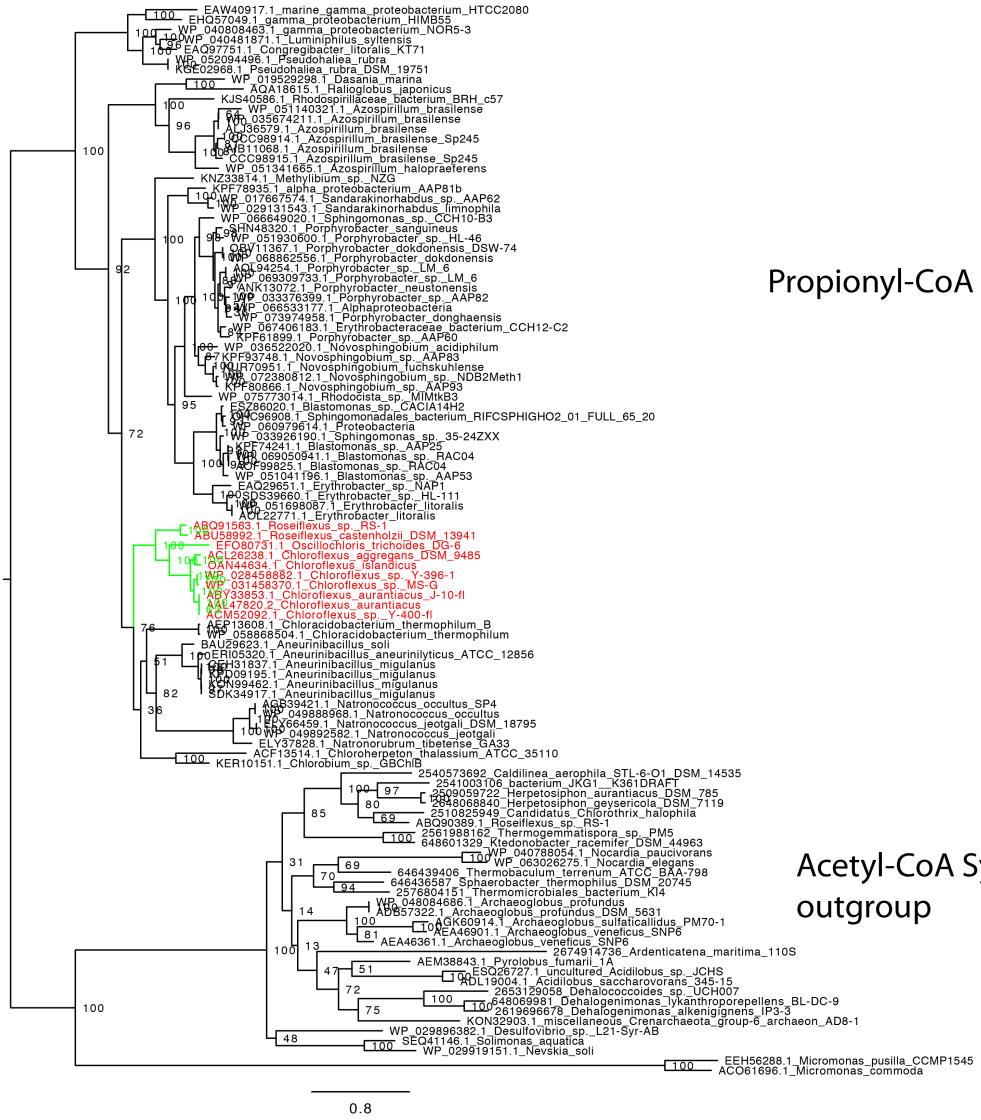

Figure S6. Maximum likelihood phylogeny of propionyl-oA synthase. Chloroflexales taxa sit in a derived position within the phylogeny, indicating that the PCS enzyme most likely did not evolve within the Chloroflexi phylum, but rather evolved in a different phylum and horizontally gene transferred into the Chloroflexales, which would have been a prerequisite event in the *de novo* evolution of the 3HP bi-cycle. The closest homologs of PCS were collected by using BLAST against the NCBI Reference Sequence Database, which were from a protein family of Acetyl Co-A Synthetase enzymes and thus could be used as an outgroup.

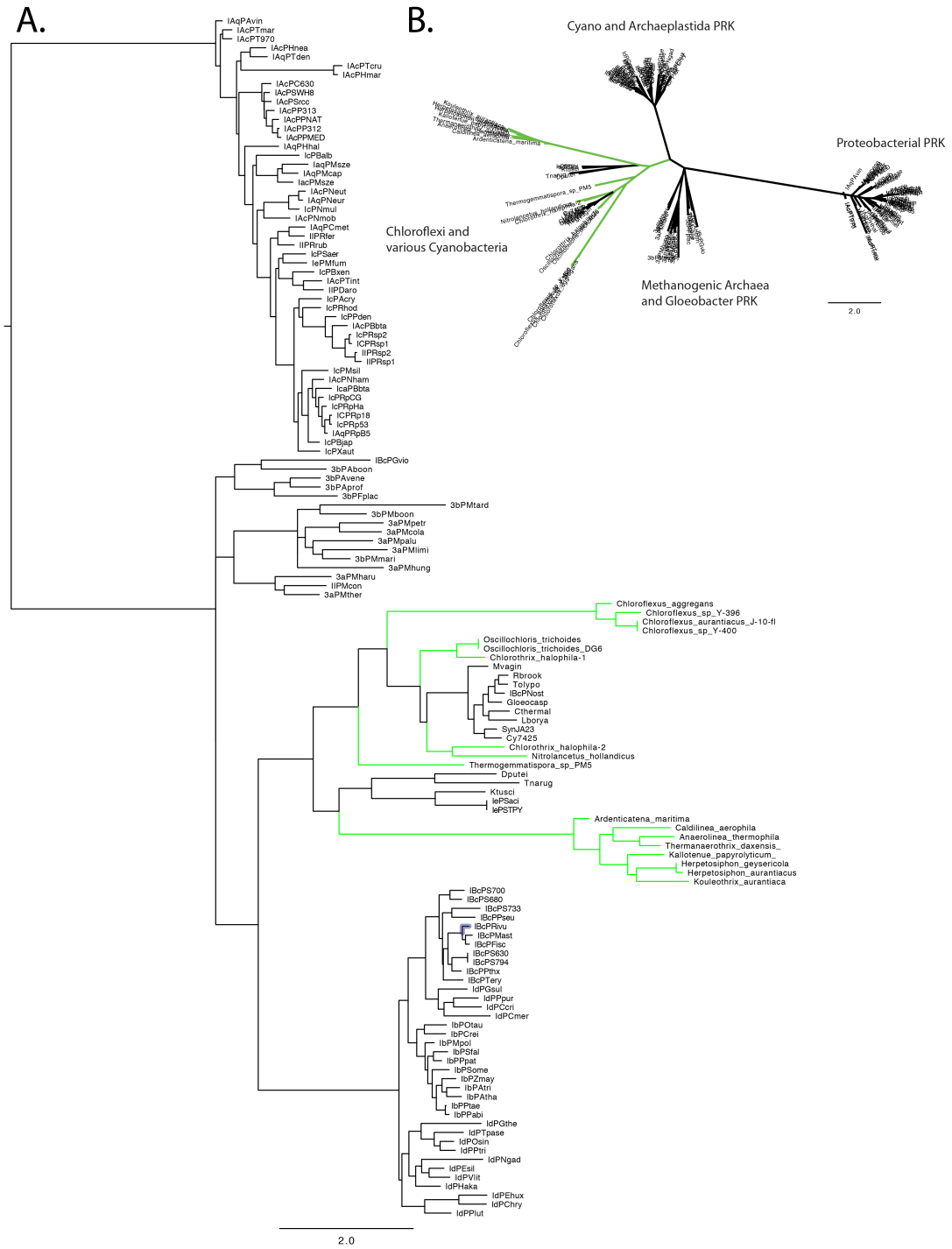

Figure S7. Maximum likelihood phylogeny of phosphoribulokinase. a, Rooted tree including a wide distribution of PRK homologs across RuBisCO containing organisms. With the addition of Chloroflexi PRK, a novel clade emerges that contains a number of PRK from Cyanobacteria, suggesting that this novel clade of PRK may have horizontally transferred into various Cyanobacteria. There is a wide distribution of Cyanobacreria within this clade, as well as a number of Form 1E RuBisCO-containing organisms. **b,** unrooted tree.

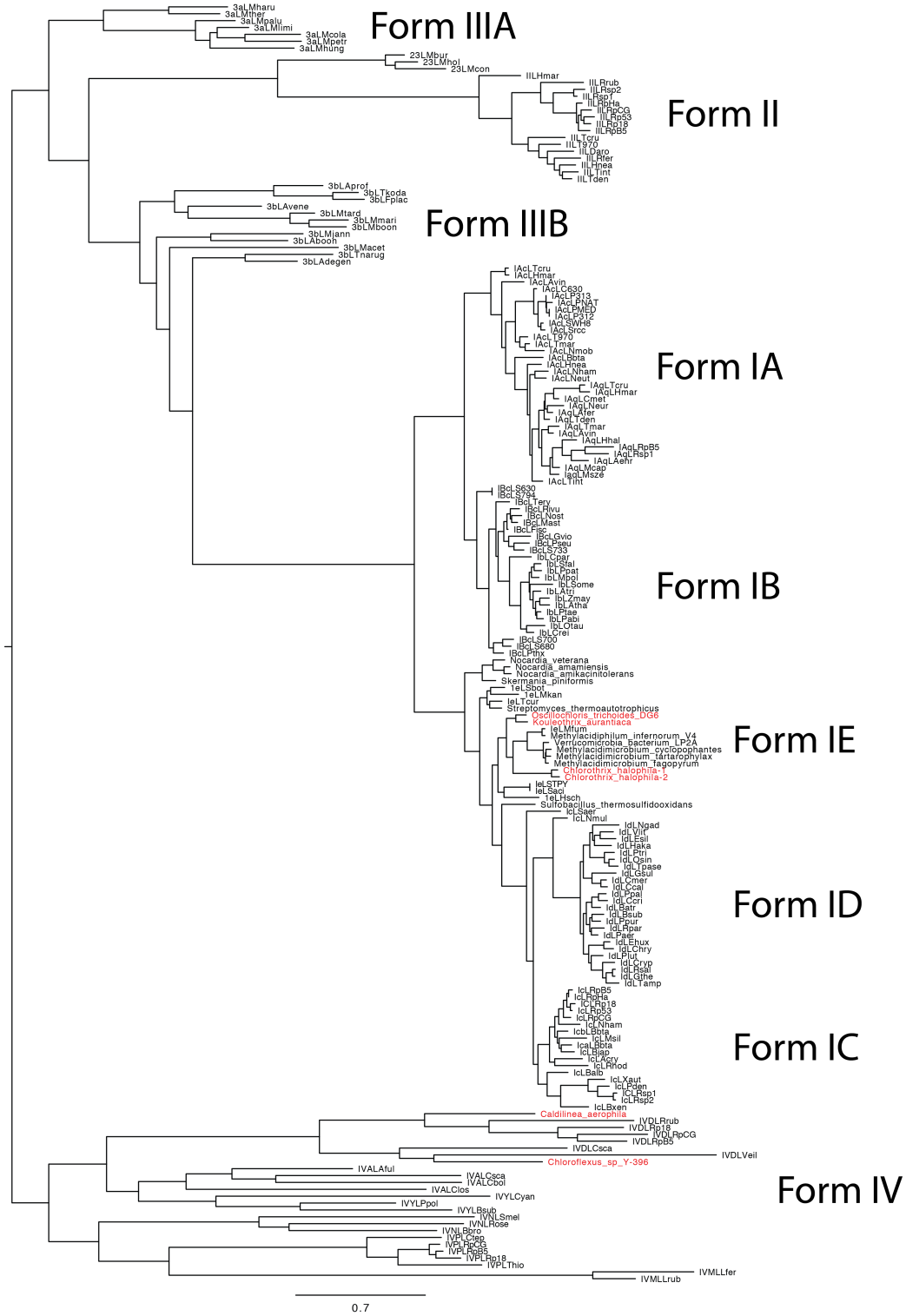

Figure S8. Maximum likelihood phylogeny of RbcL. Wide sampling of RuBisCO homologs places all Chloroflexi RbcL with in the polyphyletic Form IE clade. There are a few examples of Form IV RuBisCO in Chloroflexi; however, Form IV RuBisCO do not have carboxylase activity. All Chloroflexi taxa are highlighted in red.

Supplementary Table:

Table S1. List of taxa and abbreviations used in this study.

Abbreviation	Species	Group
DEHALK	Dehalogenimonas alkenigignens IP3-3	Chloroflexi
DEHLYK	Dehalococcoides sp. UCH007	Chloroflexi
DEHUCH	Dehalogenimonas lykanthroporepellens	Chloroflexi
THEPM5	Thermogemmatispora sp. PM5	Chloroflexi
KTEDON	Ktedonobacter racemifer DSM 44963	Chloroflexi
ATHERMO	Anaerolinea thermolimosa	Chloroflexi
LEPTOL	Leptolinea tardivitalis	Chloroflexi
LEVILI	Levilinea saccharolytica	Chloroflexi
CALDIL	Caldilinea aerophila DSM 14535	Chloroflexi
THETER	Thermobaculum terrenum YNP1	Chloroflexi
SPHAER	Sphaerobacter thermophilus DSM 20745	Chloroflexi
THEKI4	Thermomicrobiales bacterium KI4	Chloroflexi
KALLO	Kallotenue papyrolyticum JKG1	Chloroflexi
HERPAU	Herpetosiphon aurantiacus DSM785	Chloroflexi
HERPGE	Herpetosiphon geysericoa	Chloroflexi
CHL396	Chloroflexus sp. Y-396-1	Chloroflexi
CHLAGG	Chloroflexus aggregans DSM9485	Chloroflexi
CHL400	Chloroflexus sp, Y-400-fl	Chloroflexi
CHLAUR	Chloroflexus aurantiacus J-10	Chloroflexi
OSCILL	Oscillochloris trichoides DG6	Chloroflexi
CTHRX	Chlorothrix halophila	Chloroflexi
KOULEO	Kouleothrix aurantiaca	Chloroflexi
ROSCAS	Roseiflexus castenholzii DSM13941	Chloroflexi
ROSRS1	Roseiflexus sp. RS-1	Chloroflexi
AGRTU	*Agrobacterium tumefaciens* strain C58	α-proteobacteria
CAUCR	*Caulobacter crescentus* strain ATCC 19089	α-proteobacteria
RICPR	*Rickettsia prowazekii* strain, Madrid E	α-proteobacteria
RICTY	*Rickettsia typhi* strain ATCC VR-144	α-proteobacteria
mtAMBTC	*Amborella trichopoda*	Mitochondria
mtORYSJ	*Oryza sativa subsp. japonica*	Mitochondria
mtZYMAYS	*Zea mays*	Mitochondria
mtARATH	*Arabidopsis thaliana*	Mitochondria
mtPHYPA	*Physcomitrella patens subsp. patens*	Mitochondria
BP1	*Thermosynechococcus elongatus BP-1*	Cyanobacteria
PCC7116	*Rivularia sp. PCC 7116*	Cyanobacteria
PCC7120	*Nostoc sp. PCC 7120*	Cyanobacteria
PCC9605	*Fischerella sp. PCC 9605*	Cyanobacteria
PCC6803	*Synechocystis sp. PCC 6803*	Cyanobacteria

PCC7424	*Cyanothece sp. PCC 7424*	Cyanobacteria
IMS101	*Trichodesmium erythraeum IMS 101*	Cyanobacteria
PCC7367	*Pseudanabaena sp. PCC 7367*	Cyanobacteria
MED4	*Prochlorococcus marinus, subsp. pastoris CCMP 1986*	Cyanobacteria
MIT9313	*Prochlorococcus marinus MIT 9313*	Cyanobacteria
WH8102	*Synechococcus sp. WH 8102*	Cyanobacteria
PCC7942	*Synechococcus elongatus PCC 7942*	Cyanobacteria
PCC9006	*Prochlorothrix hollandica PCC 9006*	Cyanobacteria
JA33A	*Synechococcus sp. JA-3-3Ab*	Cyanobacteria
PCC7336	*Synechococcus sp. PCC 7336*	Cyanobacteria
PCC7421	*Gloeobacter violaceus PCC 7421*	Cyanobacteria
MELA1	*MEL.A1*	Melainabacteria
MELC2	*MEL.C2*	Melainabacteria
cpAMBTC	*Amborella trichopoda*	Plastid
cpARATH	*Arabidopsis thaliana*	Plastid
cpORYSJ	*Oryza sativa subsp. Japonica*	Plastid
cpZMAYS	*Zea mays*	Plastid
cpPHYPA	*Physcomitrella patens subsp. patens*	Plastid
cpCHAVU	*Chara vulgaris*	Plastid
cpCHLRE	*Chlamydomonas reinhardtii*	Plastid
cpCHLVU	*Chlorella vulgaris*	Plastid
cpCYACA	*Cyanidium caldarium*	Plastid
cpGALSU	*Galdieria sulphuraria*	Plastid
cpCYAME	*Cyanidioschyzon merolae*	Plastid
cpPORPH	*Porphyridium purpureum*	Plastid